

An Analysis of Mouse Trajectories to Ascertain Hesitation in Solving Word-Reordering Problems

Mitsumasa Zushi¹, Yoshinori Miyazaki², Takashi Miyamoto² and Ken Norizuki³

1. Department of Information Arts, Faculty of Informatics, Shizuoka University, Hamamatsu, Shizuoka Prefecture 432-8011, Japan

2. Department of Computer Science, Faculty of Informatics, Shizuoka University, Hamamatsu, Shizuoka Prefecture 432-8011, Japan

3. School of Information Studies, Shizuoka Sangyo University, Fujieda, Shizuoka Prefecture 426-8668, Japan

Abstract: We have been developing a Web application that will help both teachers and learners notice the crucial aspects of solving word-reordering problems. Evaluation systems most computers employ deal with only results of the answers reached by learners without looking into the process by which the answers are produced. It will be insufficient to ascertain learners' understanding level because correct answers may well include lucky hunches, namely accidentally correct but not confident answers. In order to differentiate these lucky answers from confident and correct ones, we have developed a Web application that can record mouse trajectories during the performance of tasks. By analyzing the recorded trajectories, our Web application will be able to identify the accidentally correct answers and notify learners of the fact that they do not fully understand the important factors tested by the problems. After a brief description of this Web application, we illustrate the way to classify confident and not confident answers as a whole in terms of mouse trajectories, and then we discuss how to apply the classification method to individual learners.

Key words: E-learning, mouse trajectory, study logs, information retrieving tool, occurrence of hesitation.

1. Introduction

A major goal of this study is to provide both learners and teachers with a practical Web application which will sensitize them to the critical aspects of solving WRPs (word-reordering problems). In WRPs learners are required to make an English sentence from given words, one with a meaning equivalent to the sentence provided in the learners' native language. This type of problem has been and is even now a popular means in Japan to measure learners' knowledge of grammatical items, sentence structure, idioms/idiomatic phrases, and usages—the command of which is essential for producing correct sentences. Since Japanese is a synthetic language, it has quite a loose word order, while in an analytic language like English, word ordering has a crucial importance to decide the meaning of the sentence, its grammaticality, and acceptability. This is why novice or poor Japanese

learners of English have difficulty in making a correct English sentence, and WRPs are effective in confirming the ability needed to produce English sentences.

In this paper, after introducing several related researches in Section 2, we first describe our Web application briefly in Section 3 and secondly illustrate the difference between confident and not confident answers in terms of some parameters for mouse movements in percentile ranks and in box-whisker-diagrams as well in Subsections 4.1 and 4.2. Thirdly, the useful parameters deeply connected to the difference will be applied to identify “hesitation” (hesitating movements of mouse caused from lack of knowledge) of individual learners in Subsection 4.3. If the personal norm of confidence could be obtained, we hypothesize, hesitation would probably be revealed in the parameters for his/her personal mouse trajectories. Lastly, the relationship between the difficulties in solving problems detected by the parameters for hesitation and the difficult

Corresponding author: Mitsumasa Zushi, professor, research fields: English education and intercultural communication. E-mail: mzushi@inf.shizuoka.ac.jp.

words learners themselves report by the use of a declaration function on the system will be discussed to examine how much they will correspond to each other in Subsection 4.4. Our future tasks are stated in “Concluding remarks” (Section 5).

2. Related Research

There have been several studies that have featured the analysis of study logs or mouse movements of learners. Focusing on the analysis of mouse trajectories, Tateda et al. [1] developed an authentication system. They illustrated a number of technical properties of individual learners’ mouse movements, including their performance speeds. Ohmori et al. [2] scrutinized mouse behavior during the course of learners’ reading task, classifying learners’ reading habits into three general patterns. Arroyo et al. [3] presented a web logging system that tracks mouse movements on websites to help web site administrators run usability tests and analyze the collected data. Ikegami [4, 5] discussed several characteristic patterns in learners’ study habits by analyzing their study logs and clarified the relationships between each study pattern and its effectiveness, and suggested effective learning environment. Freeman and Ambady [6] developed a software package called “Mouse Tracker,” which monitors subjects’ behaviors during a psychological task, with the mouse trajectories being processed and visualized in real time. Some attempts have been given to analyze facial expressions. Nakamura et al. [7] analyzed face movements in order to know the difficulty/easiness of the e-Learning problems the learners solved, and Horiguchi et al. [8] developed a system that will analyze learners’ facial responses and mouse speed so as to presume learners’ state of mind in the course of e-Learning tasks.

These studies have featured the analysis of study logs or mouse movements of learners; however, none of these studies mention the possibility of using mouse trajectory data to analyze the process by which

learners solve problems for the purpose of evaluating learners’ understanding. The foundation of our current system for measuring the understanding levels was created by Miyazaki et al. [9]. Their preliminary experiment identified several types of mouse movements which may have deep connection to hesitation in the solving process. Following their research, we have conducted several experiments with our improved software. We have analyzed the mouse trajectories, response time, and click, drag, and drop logs in the solving process, in relation to the scores and confidence levels rated by learners themselves. We discussed in Miyazaki et al. [10] and Zushi et al. [11-13], that these records would be potential indicators of the degree of learners’ understanding, which otherwise cannot be detected from mere scores.

3. Web Application

In order to achieve our goal, we have been developing software that has three modules with independent, but also interrelated functions: (1) Problem Construction Module, where teachers construct and add new problems on their own, or change marking scales from the default form; (2) Study Module, which requires learners to perform word-reordering tasks by “dragging and dropping” each word into the appropriate position in a sentence, simultaneously recording all the mouse trajectories as well as the timing of drag-and-drops (D&Ds) in answering the problems; (3) Retrieval & Analysis Module, which reproduces all the actions recorded in the learners’ mouse trajectories, and analyzes the data from the diverse patterns of the study logs both from the learners’ and problems’ perspectives. This software, consisting of three modules, has been programmed using PHP, JavaScript, and MySQL.

3.1 Problem Construction Module

This module helps teachers provide WRPs for learners to solve. The processes of making problems are shown in Fig. 1.

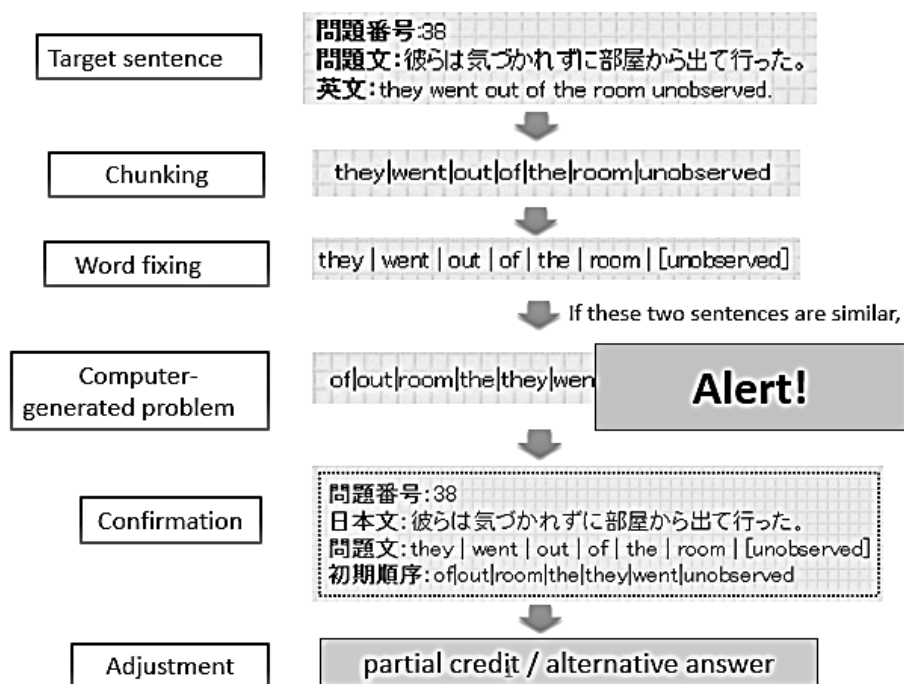


Fig. 1 Procedure for a WRP construction.

All teachers have to do is just to type correct sentences in English and the corresponding meaning in Japanese. Teachers may use “chunking” and/or “word fixing” functions if they want to make the problem easier or shorter. By using the chunking function, a series of words become one group when the straight-bar(s) “|” that separate each adjacent words are deleted, and the group of words will be treated as one word, making the number of words included in the problem smaller. The word fixing function is also available when teachers prefer to make the problem easier or shorter by literally fixing the word position in a sentence. Once a word or a group of words are fixed to a certain position, learners are not able to change it, which means that learners have only to move the rest of the words in the problem.

Furthermore, this function is the most useful when teachers wish to avoid double/multiple correct answers to one problem. The computer’s binary recognition system accepts only one generated sentence as the correct answer unless another or other correct ones have been additionally programmed in

advance. The “word fixing” function makes it easier to deal with a sentence composed of words that can be arranged into more than one correct order. For example, the sentence, “They went out of the room unobserved” can also be ordered correctly as “Unobserved they went out of the room.” Using this function, though, teachers can fix the position of “unobserved” in the initial or end position (as in Fig. 1), and thereby prevent the possibility of double correct answers.

After these procedures have been completed, teachers can choose one of three break down methods: optional, alphabetical, or randomized order. Automatic randomization is the easiest to use and will be the most preferred, but there is a rather serious danger of producing a WRP which is similar in form to the correct sentence. The similarity to the correct sentence may make the problem too easy for learners to solve. If automatic randomization produces a similar word order to the target structure, the teacher can use another break down method, an alphabetical word order, in which chances for the similarity to occur become smaller. However, alphabetical-order

arranging may not be perfect, either, creating a problem that has already formed an important structure used in the sentence. An “Alert” function is incorporated into this module to eliminate the teacher’s burden of checking these similarities. When notified by “Alert” that the problem created has a similar word order to the answer, teachers will opt to change the word order in the problem by using a third method, an optional word order, which allows the teacher to choose the order. In the last stage of making problems, confirming the target sentence (=answer) and the word arrangements in the problem, teachers can prepare partial credit for a correct part of a sentence or a sentence which is not completely correct but good enough to make sense (e.g., Olympics are watched by the people all over the world).

3.2 Study Module

In WRPs, learners are asked to reorder given words and to make an English sentence in accordance with

the sentence provided in Japanese. The words to be rearranged are given in the “problem slot,” and all the words should be moved into the “answer slot” by D&Ds with the mouse (Fig. 2). All mouse movements for solving problems are recorded here, such as D&D(s), U-turn(s) (the right-and-left or up-and-down mouse movements), and the time used in one treatment of one word, as well as the time elapsed between a particular drop and the click of the next word (D-C time), and standstill time of the mouse.

It is required for learners to press the OK button to finish answering, and then they rate their confidence level in the answers on a four-point scale. Using a newly added function to this module, they have a chance to report which word(s) have been difficult to reorder by checking the word(s) in their own answers. After all these procedures are completed, the answers are automatically evaluated by the system.

In order to facilitate learners’ performing the tasks, this module has the following functions:

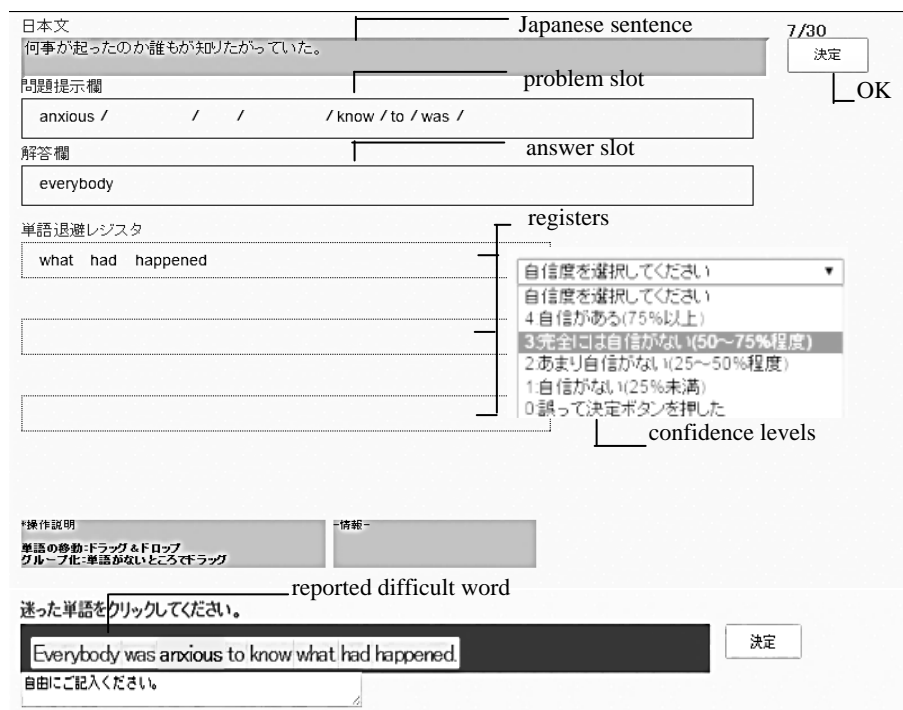


Fig. 2 Study Module.

(1) Word groupings: an arbitrary number of words can be grouped together by mouse-dragging (rectangular selection) if it is convenient for learners to treat them together, and;

(2) Relocation to registers: areas called “registers” are provided as a temporary “shelter” for words, where a set of words can be integrated into meaningful segments if it is preferred for learners to organize their ideas.

3.3 Retrieval & Analysis Module

The study log data recorded by the Study Module when learners perform tasks, can be retrieved and analyzed by using one of the six menus in the Retrieval & Analysis Module. The menus included in this module are Reproduction, Learner Analysis, Problem Analysis, Study Log Retrieval, Correlation Analysis, and Clustering menus, each of which has links to related pages of other menus so that teachers are able to have an access to sets of data they want to check (Fig. 3).

(1) Reproduction: mouse trajectories are reproduced in visible lines whose colors change every ten seconds, so that the newer lines can be identifiable when the trajectory lines overlap, with reproduction speed

changeable from 50% to 500% of its original speed, and also with reproduction being able to start at any selected point on the slider.

(2) Learner Analysis: data for a particular learner is available, such as the percentage of correct answers classified by grammatical items, the total number of problems attempted, the percentage of overall correct answers, as well as the average time needed for answering problems.

(3) Problem Analysis: data for a particular problem can be reviewed, such as the number of times the problem was attempted by all learners, the percentage of correct answers, the average time needed for answering a certain problem, the number of words to be reordered in the problem, and so on.

(4) Study Log Retrieval: this menu helps teachers retrieve data for specific criteria they want to focus on. Combining several search types and commands, the teacher can have an access to, for example, learners that used U-turns more than 15 times to solve a particular problem, or problems that required D&Ds less than 10 times for a certain learner to solve.

(5) Correlation Analysis: with this menu correlation analysis is available between scores and other parameters for mouse movements—such as response

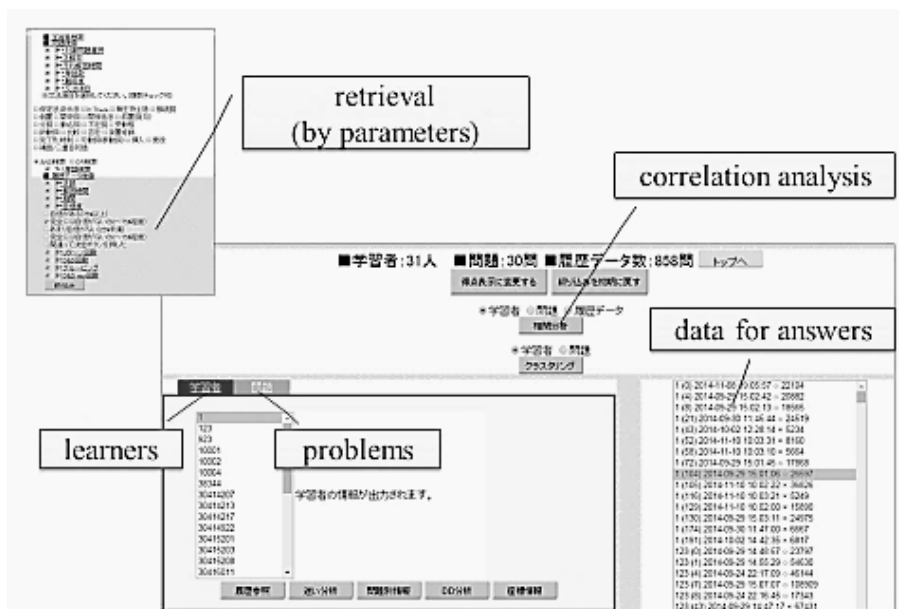


Fig. 3 Retrieval & analysis module.

time needed, the number of D&Ds, the total distance of mouse movement, and the number of U-turns—from either the learners' or the problems' perspective.

(6) Clustering: this function assembles similar logs in groups. In this study, considering the large difference in the variances and scales of each parameter, a standardized Euclidean distance was used instead of the generally used (non-standardized) Euclidean distance. Ward's method was adopted for computing the distances among clusters. Teachers can choose the number of clusters and which parameters to be incorporated for clustering criteria, from “the ratio of correct answers,” “the number of D&Ds,” “the number of U-turns,” and “average answer time.”

For more detailed description of the menus and their functions in Retrieval & Analysis Module, refer to Zushi et al. (2014a) and (2014b).

4. Experiment and Discussion

Asking 20 students with a variety of majors at a certain university in Japan, to solve 30 problems using the WRP system we have developed, we conducted an experiment in which the subjects were informed that their mouse trajectories would be recorded and analyzed for the purpose of investigating the solving processes. Five hundred and eighty-one sets of data were collected in total, with 19 sets of data being inappropriate for analysis due to a recoding failure or dropping mistakes on the part of learners.

4.1 *Confident or Not Confident*

An important consideration is how to differentiate a confident and correct answer from a lucky hunch. Both answers are correct, but one is reached by a learner who has full control of grammatical structures and lexical items to construct the right sentence, while the other answer, though correct, may be produced by a learner who possibly lacks that particular knowledge. The latter learner may have managed to reach the target structure by coincidence after a long period of

fluctuating mouse movements. The fact of being correct or wrong alone may not always be a true reflection of the ability tested unless the process of the learners' responses are carefully considered.

Learners rate their confidence in answers on a four-point scale—confident, a little confident, not so confident, or not confident—but we will limit our analysis to confident and not confident answers in order to find out typical mouse movements immediately connected to full or no confidence.

They fall into one of four categories: (1) Confident and correct; (2) Confident but incorrect; (3) Not confident but correct; (4) Not confident and incorrect. Confident answers belonging to category 1 include useful mouse movement logs for finding the norm of “answers without hesitation or uncertainty.” The answers from category 3 need very careful treatment because most computer marking systems may evaluate them as “passing” in spite of the fact that learners are not yet sure of the correct answers, or they might have reached the correct answers accidentally, with no confidence in their answers at all. There is no need of analyzing the answers of categories 2 and 4 for the purpose of ascertaining learners' understanding levels because the answers are wrong, which means that they have to review the problems again.

Mouse trajectories in Figs. 4 and 5 represent correct answers, but Fig. 4 is for an answer from category 1 and Fig. 5 from category 3. It is quite apparent that the trajectory in Fig. 4 demonstrates much simpler and smoother movement of the mouse in producing an answer than that of Fig. 5, whose trajectories overlap many times, displaying 37 times of U-turns during about 133 seconds needed for answering. Fig. 4 includes only 18 U-turns and about 37 seconds.

Because both answers are correct, the difference in mouse movements does not come from “correctness” of answers but from “confidence” in answers the learners produced.

Through the comparison of confident and not confident answers in our former experiments, it was

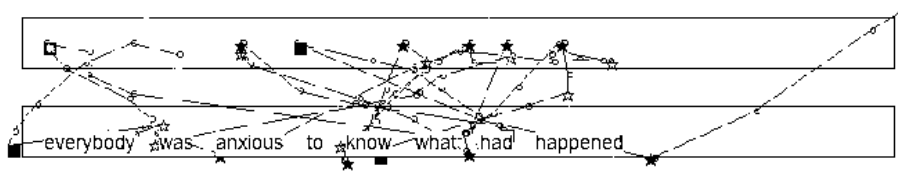


Fig. 4 Correct and confident.

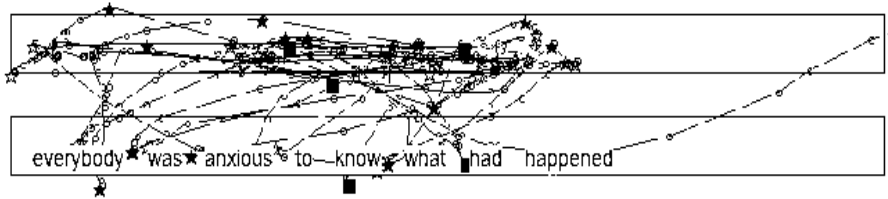


Fig. 5 Correct but not confident.

found that important parameters for mouse movements included time needed for answering, D-C time, the number of D&Ds, and the number of U-turns. Using the data collected in these experiments, we compared answers from category 1 and those from category 4, in order to investigate how confidence and parameters were related and how confidence had an effect on producing correct or incorrect answers. As is discussed in Zushi et al. (2012), answer time, D&Ds, and U-turns showed highly negative correlation with confidence, suggesting these parameters could serve as an index of hesitation resulting from learners' lack of confidence. Zushi et al. (2014b), with some more parameters such as D-C time and the longest standstill time being added, demonstrated the difference in the distribution of answers from categories 1 and 4, by investigating the deviation values of these important parameters in the answers of category 4, from the norm decided on the answers belonging to category 1. The results of these former analyses reasonably suggested that the parameters showing large differences from the norms, (especially time needed for answering, D-C time, average velocity, standstill time, and the number of U-turns), can be used as clues to detect lack of confidence and hesitation of mouse movements.

Analyses of the data from two recent experiments will further confirm the relationship between learners' confidence and their mouse movements. Table 1 is a comparison of percentile ranks of some parameters' mean values of not confident answers in these two experiments. The mean values of not confident answers are added to confident answers' data sets in order to investigate the difference (or distance) between confident and not confident answer groups. The value of 96.1 for Parameter 2 in the experiment of 2013, for example, indicates that only 3.9% of 207 confident answers (that is, approximately 8 answers) exceed the mean value of not confident answers in answer time, showing that this value is quite exceptional when calculated in the group of confident answers. And it is true of the values for other parameters.

The highness (and lowness for Parameter 1) and similarity of ranking values for each parameter in two experiments imply that these parameters provide a useful framework for understanding signs of difficulties learners experience. The farther apart these 0th to 100th percentile ranks are from the center of the scale (=50th), the less confidence and the less adequate knowledge learners may have in producing target sentences.

Table 1 Percentile ranks of category 4 (not confident and incorrect answers) mean values when they are added to category 1 (confident and correct answers) data sets.

	P1	P2	P3	P4	P5	P6
Percentile ranks	8.2	96.1	94.7	96.6	V: 96.1 H: 95.2	94.7
	7.6	100	97.5	98.1	V: 93.6 H: 98.1	96.8

*upper row(s): 2013 experiment (n=208, including the mean value of category 4)

lower row(s): 2014 experiment (n=158, including the mean value of category 4)

P1: Average velocity of mouse speed.

P2: Answer time.

P3: Total distance of mouse movement.

P4: The longest D-C time.

P5: The number of vertical (V) and horizontal (H) U-turns.

P6: The longest standstill time.

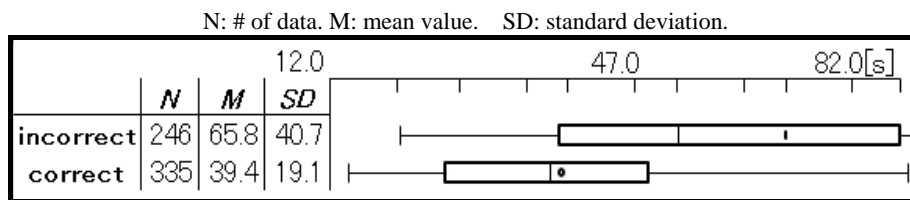


Fig. 6 Distribution of correct and incorrect answers in terms of answer time.

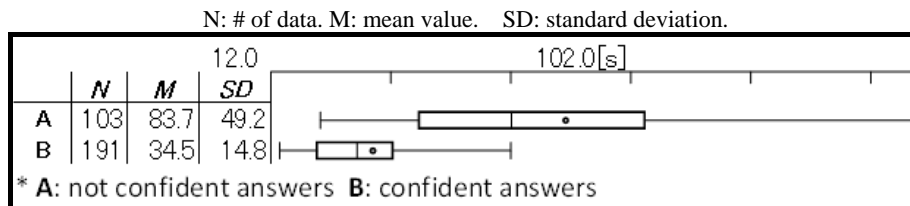


Fig. 7 Distribution of confident and not confident answers in terms of answer time.

Among these parameters, the answer time and D-C time are remarkable in both experiments. It is assumed that long answer time is connected to difficulty in solving the problem, and thus to the high rate of incorrect answers. But answer time does not necessarily separate the two groups of correct and incorrect answers clearly.

The distribution of 335 correct and 246 incorrect answers in terms of answer time is illustrated in a box-whisker diagram (Fig. 6). A box-whisker diagram is an effective way to compare two or more groups of data sets' distribution, because it displays visually how distant apart (or overlapping) the distributions of data are. The wider the distance between the boxes is, the larger difference in value the groups of data indicate.

Answer time, as seen in the diagram, is not a reasonable classifier of correct and incorrect answers. The middle range of distribution is a domain both types of answers share. Compare this with the diagram (Fig. 7). The two boxes in Fig. 7 include no overlapping area, with a small gap between them, demonstrating clear separation of confident and not confident groups of answers in terms of the time needed for answering.

It is obvious from this comparison of two diagrams that answer time is a better indicator of learners' confidence, and thus more deeply connected to adequacy of learners' knowledge and ability tested in WRPs.

But the samples analyzed in Fig. 7 include all confident and not confident answers, namely all the

answers belonging to one of 4 categories (confident-correct, confident-incorrect, not confident-correct, and not confident-incorrect). What is important for teachers to know is how to discriminate confident and correct answers from not confident but accidentally correct answers resulted from lucky hunches.

4.2 Correct Answers and Confidence

We will now limit our investigation only to correct answers, and examine if answer time can be a classifier for the correct answers which include both confident and not confident answers. Fig. 8 illustrates the distribution of confident-correct answers and not confident-correct answers in terms of answer time.

Both types of answers are correct, but they are far apart from each other in terms of confidence levels, and the difference between confidence levels is manifested in the time difference needed for answering. Thus, time for answering and learners' confidence are deeply connected, showing this parameter as a clue to uncertainty or hesitation in the process of solving problems. If an e-Learning system detects problems whose answers are correct but their trajectories belong to category 3 (not confident-correct answers), the system should not pass the learners but demand them to review more of the same type of problems.

Another parameter whose value is among the highest as well as the answer time in Table 1, is the longest D-C time, and it also can be an indicator for confidence as shown below (Fig. 9). The more uncertain the learner is of the correct sentence, the longer the D-C time will be somewhere in performing a task. This time can be interpreted as hesitation of mouse movements, because if the learner has clear word order in his/her mind, a long time would not be elapsed to look for the next word to move after a drop of a certain word.

Now that these parameters have been verified as useful difficulty indices for a group of students as a whole, we need to further examine if the same parameters will be applicable to personal analysis. Some learners who may be careful in D&Ding words, or other learners who are not accustomed to mouse control so much, may need longer answer time. And the result would be that the system may judge the slow mouse movements are caused from hesitation or uncertainty on the part of learners even if they are very certain and confident of the correctness of sentences they are asked to produce.

4.3 Confident and Not Confident Answers of Individual Learners

Each learner solved only 30 problems, so the number of "confident and correct" or "not confident

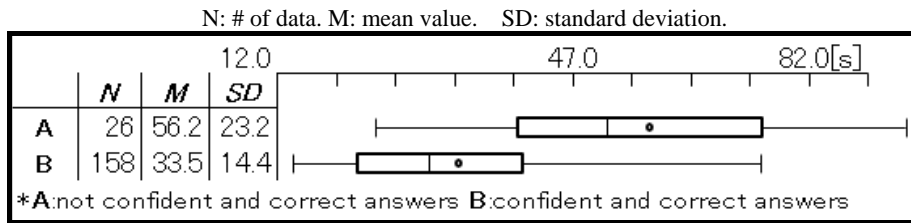


Fig. 8 Distribution of "confident-correct," and "not confident-correct" answers in terms of answer time.

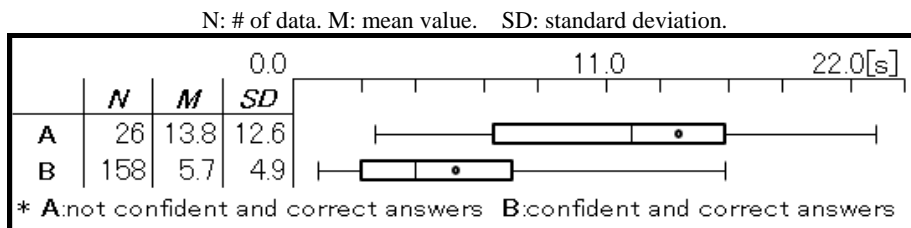


Fig. 9 Distribution of "confident - correct," and "not confident-correct" answers in terms of the longest D-C time.

but correct” answers collected from the experiment is not large enough to be treated mathematically in a diagram or by deviation values. But it may be worth comparing the individual raw values derived from the useful parameters which have been found in group analysis.

Table 2 is a very typical sample, in which the learner has 6 confident and correct answers, and 7 not confident but correct answers. All parameters, especially P2, P4 and P5, indicate that the mouse movements from two groups of answers are quite different. Not confident answers, in average, need about 10 seconds more for answering, about 3 seconds longer for deciding the next word to move, and use about 10 more U-turns than confident answers.

Another learner, whose answers illustrate clear difference of parameters, is a less typical but good enough example to know that these parameters can be classifiers for confident and not confident answers on an individual basis (Table 3).

These values are calculated from 11 answers, 8 of

which are confident and correct, and 3 are not confident but correct answers. Here again, the difference between confident and not confident answers is remarkable in terms of P2, P4 and P5.

If each learner solved many more problems than 30, it would be possible to decide personal norms for each of individual learners based upon his/her confident and correct answers. And if the personal norm could be determined, correct but not confident answers would be detected by treating their parameters mathematically.

4.4 Reported Difficulties and Parameters

Self-declaration function, a new function added in the Study Module, reports the difficult words or phrases learners themselves have experienced while performing tasks. If the reported difficulties include the parameters connected to hesitation, we hypothesize, these parameters would be valid identifiers to ascertain when and where the hesitation occurs in the process of solving one problem.

Table 2 Mean values of important parameters: a comparison of confident and correct answers with not confident but correct answers—learner A.

	P1	P2	P3	P4	P5	P6
confident and correct	0.19	30.50	5,314	4.67	19.33	2.96
not confident and correct	0.21	41.31	7,438	7.26	28.71	3.30

P1: Average velocity of mouse speed (px/ms).

P2: Answer time (s).

P3: Total distance of mouse movement (px).

P4: The longest D-C time (s).

P5: The number of U-turns.

P6: The longest standstill time (s).

Table 3 Mean values of important parameters: a comparison of confident and correct answers with not confident but correct answers—learner B.

	P1	P2	P3	P4	P5	P6
confident and correct	0.22	33.13	6,838	4.81	29.25	2.33
not confident and correct	0.17	66.29	10,590	12.45	48.00	2.54

P1: Average velocity of mouse speed (px/ms).

P2: Answer time (s).

P3: Total distance of mouse movement (px).

P4: The longest D-C time (s).

P5: The number of U-turns.

P6: The longest standstill time (s).

We investigated occurrences of D-Cs (counted as an outlier when its time exceeds the rejection region fixed as 1) and the use of extra D&Ds (D&Ds used more than necessary in solving a problem), assuming they would be good locators to learners' hesitation in performing a task. Among 158 correct and confident answers, the reported difficult words/phrases are only 8 (learners are confident of their answers and have few difficult experiences in producing correct sentences). But the words/phrases including one of the two parameters we selected, sum up to as many as 89. Even among 26 correct but not confident answers, declared difficulties are 12, but the parameters suggest 29 difficult words/phrases. Though the parameters respond better to not confident answers, and though the parameters successfully detect difficult problems, they cannot always be effective locators which help us find where in a problem or which word/phrase in a problem is difficult for a certain learner.

The result of the experiment requires drastic adjustment of criteria for parameters or alternation of parameters with which learners' uncertainty or hesitation is identified, because they have a strong tendency to apprehend many unrelated mouse movements to declared hesitation. Slowdown of some mouse movements leading to a long D-C time, which our system has picked up as signs for hesitation, may be caused by learners' inadequate understanding of the overall structure of the target sentence, and irrelevant to learners' uncertainty of a particular word's position. Also, extra D&Ds may result from accidentally dropping a word into a wrong position, and it will require the learner to replace the word into an appropriate position with another D&D. These elements need to be excluded from the count of hesitating parameters.

Answer time, which constantly shows high conformity to uncertainty or hesitation, cannot be utilized for the purpose of locating which part of a problem is difficult for learners. But it would be worth examining other parameters such as U-turns, D-D time

(time for one D&D), and their combination with D-C time and/or extra D&Ds, because they reveal stronger connection to not confident answers than confident ones as well as the above two parameters.

5. Concluding Remarks

Parameters found useful in our earlier analyses have been shown in the latest experiment and analysis again to be valid in detecting difficult problems for the group of learners, and the same parameters will be applied to find not confident answers reached by individual learners. But these parameters are not immediate locators to the occurrence of hesitation in the process of solving one problem, partly because of inappropriate criteria for the parameters that indicate hesitating mouse movements, and partly because of disagreement between slowdown of mouse speed and learners' uncertainty of the position of a word being dragged. Our future tasks, therefore, should include:

- (1) Adjusting criteria for D&Ds and D-C time based on which the parameters are to judge whether the mouse movements include hesitation or not;
- (2) Investigating other parameters such as U-turns, D-D time, standstill time, or combination among these, or combination with other parameters, if they can be useful clues to hesitation, and;
- (3) Conducting experiments where each learner solves far more than 30 problems, so that confident and correct answers can be mathematically dealt with as the norm for the mouse movements of the learner.

In order to fulfill these tasks, it is required to improve our system in terms both of teachers' analysis and learners' use.

Acknowledgment

This work was supported by a Grant-in-Aid (KAKENHI) for "Scientific Research (C) (25370621)", and we also owe special thanks to Naho Atsumi (a student from Faculty of Informatics, Shizuoka University) for her help with the data processing indispensable in our study.

References

- [1] Tateda, C., Kotani, K., and Horii, K. 2005. "Evaluation of an Individual Authentication System Using the Trajectory by the Mouse." *IEICE Technical Report, HCS* 105: 17-20.
- [2] Ohmori, Y., Iibushi, D., Horie, I., and Itoh, K. 2006. "Understanding Learning Effect Using MBC (1)-Analysis of Mouse Behavior." Presented at the Conference on IT Education.
- [3] Arroyo, E., Selker, T., and Wei, W. 2006. "Usability Tool for Analysis of Web Designs Using Mouse Tracks." In *Proceedings of CHI EA '06, ACM*, 484-9.
- [4] Ikegami, M. 2006. "Enhancing Students' English Reading and Listening Skills through the Use of the Computer-assisted Intensive English Training Program." *Journal of Language and Literature* 26: 103-26.
- [5] Ikegami, M. 2009. "Enhancing Students' English Reading and Listening Skills through the Use of the Computer-assisted Intensive English Training Program III—The Effect of Learning Environment and Course Design." *Journal of Language and Literature* 29: 229-57.
- [6] Freeman, J. B., and Ambady, N. 2010. "Mouse Tracker: Software for Studying Real-Time Mental Processing Using a Computer Mouse-tracking Method." *Behavior Research Methods* 42: 226-41.
- [7] Nakamura, K., Kakusho, K., Murakami, M., and Minoh, M. 2010. "Estimating Learners' Subjective Impressions of the Difficulty of Course Materials by Observing Their Faces in E-Learning." *Journal of the Institute of Electronics, Information and Communication Engineers* 93: 568-78.
- [8] Horiguchi, Y., Kojima, K., and Matsui, T. 2010. "An Automatic Estimation System of Learners' Impasses Based on Features in Low-Level Interactions in an E-Learning Environment." Presented at The 24th Annual Conference of the Japanese Society for Artificial Intelligence.
- [9] Miyazaki, Y., Tsukuda, T., and Uegaki, T. 2010. "Software Development for Recording Learners' Mouse Trajectories with Retrieving and Analyzing Functions of their Study Logs." *Journal of Japan E-Learning Association* 10: 21-30.
- [10] Miyazaki, Y., Zushi, M., and Norizuki, K. 2011. "Web Application for Recording Learners' Mouse Trajectories and Retrieving Their Study Logs for Data Analysis." In eds. J. Cheng, G. Fu, and X. Han, *Technology Enhanced Language Learning*: 95-9.
- [11] Zushi, M., Miyazaki, Y., and Norizuki, K. 2012. "Web Application for Recording Learners' Mouse Trajectories and Retrieving Their Study Logs for Data Analysis." *APTEL 2011 Best Papers: Knowledge Management & E-Learning: An International Journal* 4: 37-50.
- [12] Zushi, M., Miyazaki, Y., and Norizuki, K. 2014a. "Web Application for Measuring Learners' Knowledge of English Syntax: Analysing the Relationship between Mouse Trajectories and Learners' Understanding." *Retrospect and Prospect: Essays from the 2013 Taiwan e-Learning and Digital Archives Program International Conference*, A special supplement of the *International Journal of Humanities and Art Computing (IJHAC)*, Edinburgh: Edinburgh University Press, Vol. 8: 190-205.
- [13] Zushi, M., Miyazaki, Y., Miki, R., and Norizuki, K. 2014b. "Development of a Web Application: Recording Learners' Mouse Trajectories and Retrieving Their Study Logs to Identify the Occurrence of Hesitation in Solving Word-Reordering Problems." *Journal on Systemics, Cybernetics and Informatics: JSCI* 12: 86-91.